

## 3.1 Allgemeine Angaben zum Teilprojekt A01

### 3.1.1 Titel:

The Syntactic Expression of Information Structure and the Architecture of Grammar

### 3.1.2 Fachgebiete und Arbeitsrichtung:

General Linguistics, Syntax

### 3.1.3 Projektleitung

Fanselow, Gisbert, Prof. Dr.

Vicente, Luis, Dr.

Department Linguistik, Universität Potsdam, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam

## 3.2 Zusammenfassung

The overall goal of A1 for the third phase of the SFB 632 lies in the formulation of a model for the interaction of narrow syntax and information structure. Much recent and current work on this topic assumes the base hypothesis of the so-called cartographic approach (Rizzi 1997 *et seq*), where information-structural notions like topic and focus are directly represented in narrow syntax as formal features heading functional projections. The cartographic approach predicts, by design, that the effects of information structure on syntax are categorical (i.e., factor *F* invariably triggers structure *S*). However, available evidence (Keller and Alexopoulou 2001 and others) indicates otherwise –i.e., information structure effects on syntax are optional/gradient (factor *F* triggers structure *S* only in a certain percentage of cases). In order to account for the observed data distribution, we follow a model (cf. Fanselow and Lenertová 2010) where narrow-syntactic operations are triggered exclusively by morphosyntactic features and therefore cannot be influenced by information-structural factors (section 3.4.1.1). We refer to this as the *Strong Modularity Hypothesis* (SMH).

The approach we advocate raises the question of how to deal with the influences of information structure on syntactic structure (e.g., focus and topic movement, among others). Part of the project will consist of the development of an Optimality Theoretic model of grammar where the output of core syntax (a set of candidate structures) is inputted to an evaluation component. This component contains, among others, a number of constraints linking narrow-syntactic configurations and information-structural factors, which allow us to determine the degree of markedness of specific structures given specific contexts. Consequently, the influence of information structure on narrow syntax falls out from the fact that these constraints will rank some candidate structures higher (i.e., less pragmatically marked) than others. Additionally, we will explore OT models incorporating weighted constraints (cf. section 3.4.1.2) in order to account for the gradience and the optionality inherent to the syntactic expression of information structure.

In order to support the theoretical work just defined, we are planning a number of empirical studies (both acceptability rating and production tasks) aimed at precisely determining the strength of the different pragmatic factors involved. Note that part of the goal of A1 is to ascertain the adequacy of the SMH in languages with prosodic and morphological systems different from those investigated during the second phase of the SFB 632. As a consequence, we plan to carry out these experiments on a number of different languages – see section 3.4.1.3.

## 3.3 Bisherige Entwicklung des Teilprojekts

### 3.3.1 Bericht

In the current funding period, the stated goal of A1 was “the identification of an optimal model for the interaction of syntax and phonology in the context of the expression of information structure”. Three subgoals were

defined, *viz.* the phonology syntax interaction, prosodic correlates of information structure and recursion of information structure.

### ***Syntax-prosody interaction (to be continued)***

In the first phase, Fanselow (2004) attempted to account for the intervention effects exerted by structurally accented phrases on the movement of foci and parts of foci to the left periphery in German by assuming that prosodic features figure in the syntactic derivation, too. Our work in the second phase led to a significant improvement of these early attempts. Fanselow & Lenertová (2010) derive the prosodic intervention effects in German, Czech, and other languages in terms of a model that ties the assignment of structural accents to the process of cyclic linearization in the sense of Fox & Pesetsky (2005). It is postulated that structural accents are assigned immediately when two syntactic objects are merged together in a syntactic derivation, and it is assumed that this process presupposes the prior linearization of the two syntactic objects involved. The resulting theory also accounts for data that show that movement to the left periphery is not triggered by syntactic features related to information structure, and the authors present evidence that the model can also be successfully applied to languages such as Italian and Hungarian that figured prominently in proposals assuming a direct representation of information structure concepts in syntax. Fanselow & Lenertová (2010) conclude that distinctions of information structure play no direct role in the syntactic computation, as envisaged in Chomsky (2008). They are only indirectly relevant, insofar as they influence the prosodic organization of a clause that interacts with linearization. These ideas are in harmony with recent trends in the literature (e.g., Horvath's (2010) work on Hungarian), and they fall in line with results of project **A5** (e.g., Hartmann & Zimmermann, submitted) and **C1**.

The task of extending the model to syntactic constructions different from left peripheral movement was taken up in Fanselow (accepted), a paper that shows that German scrambling, too, is amenable to a treatment avoiding the direct syntactic representation of information structure. A particular challenge in the analysis of scrambling lies in the fact that 'normal' word order, be it derived by movement steps or not, comes with no scope ambiguity for quantifiers. Fanselow (accepted) argues that linearization statements in the sense of Fox and Pesetsky are not only linked intrinsically to stress assignment but also to the creation of scope statements regulating the semantic interaction between phrases in a clause.

In yet unpublished work presented at the Recursion conference in Amherst, Mass in 2009, Ishihara extends such ideas by applying the Fox-Pesetsky model to prosodic prominence. A prominence statement "A is more prominent than B" created cyclically in the derivation must be obeyed at any later stage, and hence sets restrictions on the prosodic realization of B. The cyclic nature of the model accounts for prosodic recursion. Correlations between restrictions on word order and relative prominence found in three typologically unrelated languages (German, Japanese, and Hungarian) further support this analysis. The neat integration of the details of Ishihara's and Fanselow's proposals remains on the agenda of project A1.

Ishihara (2009) explores an OT-account of the Condition on Copy Deletion proposed by Trinh (2009). Movement is often conceived of as a sequence of copying and deletion. One argument for this view lies in the observation that words or phrases are sometimes repeated in constellations that would otherwise simply be realized as movement. The deletion part of the movement sequence seems to not have applied in such cases. Trinh proposes a syntactic condition for the (non-) application of the deletion part of movement. Ishihara suggests that such operations are regulated by prosodic constraints. In this analysis, a (allegedly) syntactic condition on copy deletion is reinterpreted in terms of interaction between prosodic constraints.

Féry (submitted b) adds a new dimension to the prosody-syntax interaction by proposing an original approach to the problem of the universal prosodic correlate of focus, which has been traditionally claimed to be prominence (Jackendoff 1972, Truckenbrodt 1995, Büring 2009). She shows that prominence, understood as a pitch accent, is an optional correlate of a more general phenomenon, *viz.* alignment of a focus with a higher prosodic constituent. The absence of prosodic prominence in the focus structure of numerous languages has been one of the main results of the first two phases of the SFB. In this paper, Féry surveys some of these languages as well as the literature, and shows that an optimality-theoretic constraint Align-Focus is making very good predictions as a universal tendency instead of pitch accent. Being an OT constraint, Align-Focus is violable and can be dominated by higher constraints.

### ***Prosodic correlates of information structure (not to be continued)***

The primary subgoal of the project was to disentangle the effects of syntax and semantics on prosody. The assumption pursued in the last four years (see, e.g., Féry & Ishihara (2009, 2010), on Second Occurrence Focus and on the prosody-syntax interactions in Japanese and German) was that syntax has a main effect on phrasing, and semantics mostly affects the height of pitch accents which are higher (or corresponding to higher reference lines of phrases) when a focus is present, and lower (or corresponding to lower reference lines of phrases) under givenness. The intention was also to include typological data, in order to test this assumption in a cross-linguistic perspective.

With this background in mind, Féry (2010, 2011) investigated the impact that German discourse particles have on prosody. Following suggestions by Beck (2006), Eckardt (2001) and others, she proposed that

particles which change their meaning under stress, like *wieder*, *selbst*, *schon* and to a lesser extent *auch* can be accounted for by a single analysis. They are either focus particles, in which case they associate with a focused element, or they are focused themselves, in which case they possess correlates of focus, like stress, the faculty of triggering a set of alternatives, and a different syntax.

Ishihara (submitted b) examines prosodic effects of focus and syntactic structure in Japanese in a phonetic experiment. The results show that contrary to previous claims (Pierrehumbert & Beckman 1988, Selkirk & Tateishi 1991, Truckenbrodt 1995) focus does not necessarily insert a prosodic phrase boundary, although it expands the pitch range of the focused word (a potential case where Align-Focus proposed by Féry (submitted b) is outranked by other constraints). At a syntactic XP-boundary, though, a prosodic phrase boundary was consistently observed. This result opened up a new theoretical question in terms of prosodic effect of focus. Ishihara (submitted a) proposes a new analysis of prosodic effects of focus in Japanese, in which the principles of syntax-prosody mapping and the prosodic effects of focus are clearly distinguished.

The study of syntax-prosody mapping and prosodic effects of focus was extended to Hungarian. In collaboration with Balázs Surányi and Barbara Ürögdi (Hungarian Academy of Sciences), a series of phonetic experiments have been conducted. Ishihara, Surányi & Schubö (submitted) examine the prosodic partition of Topic-Comment structure in Hungarian. In this study, an experiment was conducted to investigate the prosodic behavior of moved quantifier phrases, which are syntactically placed in the position between topics and the preverbal focus position. The inter-speaker variation found in the results shows an interesting interaction between the Syntax-Prosody Mapping principles and the Stress-Focus Correspondence principle of Reinhart (1995). Ishihara & Ürögdi (accepted) investigated the prosody of complement clauses in Hungarian. The experimental results show that semantic factivity of the main verb does not exhibit significant effects on f0-realization and does not correlate with givenness, while syntactic encoding of referentiality plays a role.

Finally, Feldhausen (2010) published his dissertation, written in the framework of the project, on the interaction between syntax, prosody and information structure in Catalan

### **Recursion of Information Structure (not to be continued)**

We had started research on the question of whether the informational organization of a sentence resembles grammatical structures in being recursive in the first funding period. Our work relied on the assumption of recursion of the prosodic hierarchy (Féry 2009). Féry & Schubö (2010) investigated the recursion of center-embedded relative clauses in German and Hindi, and showed that German prosody is recursive in the same way as syntax, but Hindi prosody is not. The same result was achieved in another experiment by Féry & Kentner (2010) which reproduced an experiment by Wagner (2005) on grouping of names. This is a preliminary result. but it opens up interesting hypotheses on typological aspects of intonation. In his MA-thesis, Schubö (2010) further built on the German part of this study and investigated recursive prosodic structure of complex sentences in German. The results confirm the presence of syntax-triggered prosodic recursion.

Féry (submitted a) studies prosody and sentence accents in German, and proposes that the integration of a predicate and an argument is a larger phenomenon than assumed until now. She proposes that the head of any prosodic phrase can be an embedded prosodic phrase, and that the resulting structure is recursive. One of the consequences is that integration is to be understood as a phonological phenomenon, based on syntax, and not as a purely syntactic phenomenon, as assumed until now (see for instance Cinque 1993).

Ishihara (submitted b) showed that downstep in Japanese never shows a complete reset even when it is expected from the syntax-prosody mapping. This fact is taken to be evidence that prosodic phrasing is recursive in Japanese, and accordingly, downstep also takes place in a recursive fashion. Ishihara & Ürögdi (accepted) showed that a non-referential complement clause (which is claimed to be an embedded speech act) is realized with a more matrix-like prosody than a referential complement. This suggests that recursion of speech act is possible under certain circumstances, and that information structure could be recursively expressed (e.g., a topic in an embedded clause).

## **3.3.2 Aus dem Teilprojekt entstandene Publikationen**

### (a) Publikationen mit peer reviewing

#### (i) Internationale Zeitschriftenpublikationen

1. Fanselow, G. & D. Lenertová (2010). Left peripheral focus: Mismatches between syntax and information structure. *Natural Language & Linguistic Theory* (online first: DOI 10.1007/s11049-010-9109-x).
2. Féry C. & F. Schubö (2010). Hierarchical prosodic structures in the intonation of center-embedded relative clauses. *The Linguistic Review* 27, 289–313.
3. Ishihara, Shinichiro. 2009. Interpreting the CCD prosodically. *Theoretical Linguistics* 35:277–287.

#### (ii) Sonstige referierte Publikationen

4. Fanselow, G. (accepted). Scrambling as Formal Movement. Angenommen. In I. Kučerová & A. Neeleman (eds.). *Information Structure: Contrasts and Positions*. Cambridge, CUP.

5. Féry, C. (2009). Syntax, Information Structure, Embedded Prosodic Phrasing and the Relational Scaling of Pitch Accents. In N. Erteschick-Shir & L. Rochman (eds.), *The Sound Patterns of Syntax*. Oxford University Press. 271–290.
6. Ishihara, S. & B. Ürögdi (accepted). The syntax-prosody interface and sentential complementation in Hungarian. In T. Laczko, C. Ringen and G. Rákosi (eds.), *Approaches to Hungarian* Vol. 12, New York: Benjamins.
- (b) andere Publikationen
7. Féry, C. (2010). Information Structure of *schon*. In G. Fanselow & T. Hanneforth (eds.), *Language and Logos. Festschrift für Peter Staudacher*. 158-173.
8. Féry, C. (2011). Prosody and information structure of the German particles *selbst*, *wieder* and *auch*. In T. Borowsky et al. (eds.), *Prosody Matters: Essays in Honor of Elisabeth O. Selkirk*. London: Equinox Press. <http://www.equinoxpub.com/equinox/books/showbook.asp?bkid=417>
9. Féry, C. & G. Kentner (2010). The prosody of grouping names in German and Hindi. *Proceedings of Speech Prosody 2010* in Chicago. url: <http://speechprosody2010.illinois.edu/papers/100014.pdf>

## 3.4 Planung des Teilprojekts

### 3.4.1 Ziele

One of the general goals of the SFB 632 is to explore the general architecture of the interactions between grammar and information structure. In the upcoming phase, A1 will be working towards the fulfilment of a specific aspect of this general goal, *viz.*, exploring and modelling the way in which information structure affects constituent order and triggers syntactic operations. More specifically, we assume that the core syntactic component of grammar (narrow syntax) operates independently of information structure factors (*pace* Chomsky 2008, Fanselow and Lenertová 2010, Horvath 2010, and *contra* Rizzi 1997, Aboh 2008). Any apparent effects of information structure on syntax are mediated indirectly: prosody and semantics play a role for syntax, and they themselves respond to information structure. Furthermore, a post-syntactic Optimality Theoretic evaluation component defines the degree of markedness of specific syntactic structures in specific pragmatic contexts.

We believe that this hypothesis (which we refer to as the *Strong Modularity Hypothesis*, SMH, after Horvath 2010) defines a promising framework for the description and explanation of well-known but often-ignored non-categorical effects associated to the syntactic expression of information structure –in particular, *optionality* and *gradience*. Within this general plan of attack, we can identify three different but inter-related subgoals, which we define below.

#### **Goal 1: Core syntax and the architecture of grammar**

The SMH predicts, by design, that the notions of focus and topic are not encoded morphosyntactically (i.e., there exist no [+FOCUS] or [+TOPIC] formal features). This prediction was already successfully tested by Fanselow and Lenertová (2010) for a range of A-bar dependencies in various languages. We will follow up on these results by attempting to extend the SMH to other empirical domains, *viz.* A-movement, the morphological and prosodic expression of Information Structure, and association with focus. We will complement this work by exploring the formal underpinnings of the SMH, i.e., the way in which the relevant sound-meaning correspondences can arise in the absence of morphosyntactic encoding.

In order to ensure the success of this Goal, we plan to offer visiting professorships to two scholars with extensive expertise on these topics, namely, Luigi Rizzi (University of Siena) and Caroline Féry (University of Frankfurt). Rizzi is one of the leading proponents of the view that information-structural notions correspond to narrow-syntactic formal features, therefore his input will be crucial to identify problematic points in our SMH-based analyses. Féry has worked extensively on prosody, therefore her collaboration will enable us to specify the relation between narrow syntax and its prosodic realization in terms of phrasing constraints rather than feature-driven prominence specifications.

#### **Goal 2: Modelling gradience and optionality**

The second goal involves the development of a model that can account for the inherent optionality and gradience of syntax-information structure interactions. Our work here will revolve around Optimality Theoretic models with weighted constraints, given that these are especially well-suited to deal with optional and gradient data (cf. Keller 2000, Scholz and Pullum 2003, Jäger and Rosenbach 2006). In addition to the pure data-explanation aspect of this research, we will also aim at providing a better understanding of the nature of the OT evaluation component *vis-à-vis* the overall architecture of grammar. Does it reflect a general division of labor in syntax (Pesetsky 1998) or is it related to the core vs. periphery distinction (Barbiers 2005)? Are there domains beyond information structure to which an evaluation component applies non-trivially?

While very theory-oriented, this Goal requires a close collaboration with the empirically-oriented research of the SFB. In particular, the analysis of gradience will feed off the acceptability judgement data outputted by Goal 3 (see below). The analysis of optionality will involve a collaboration with **A6**, given that optionality is sometimes defined on the basis of corpus frequency data; therefore, we envision a joint effort to investigate possible connections between relative acceptability and corpus frequency (*pace* Bader 2010).

### **Goal 3: The establishment of the empirical basis for modelling.**

In order to establish the empirical basis for Goal 2, we plan a set of acceptability rating and production experiments, to be run in parallel. The factors under investigation involve “local” aspects of the syntax-information structure interface (e.g., the effect of corrective focus on word order) as well as “global” aspects that relate to text structure (e.g., the preparation of a topic shift) or the communicative situation (e.g., attention control or emotional states). These experiments will shed light on the question of to what extent the actual choice of a specific syntactic structure is determined by information structure in the classical sense of the term. More broadly, the results of these experiments will be collected in a first-of-its-kind database that will be of great use to all researchers (both internal and external to the SFB 632) interested in these issues.

## **3.4.2 Work Program**

### **3.4.2.1 Goal 1: Core syntax and the architecture of grammar**

There is a tradition in syntax holding that concepts of information structure systematically correspond to certain syntactic categories, such as heads and their projections (Rizzi 1997, Aboh 2008) or formal features (Brody 1995, Bošković 2002). On the other hand, there are approaches, sided by our previous work in project A1 and other projects of the SFB, which explicitly deny the existence of any discourse-related categories in syntax (cf. Chomsky 2008, Neeleman & van de Koot 2008, Horváth 2010, Fanselow and Lenertová 2010). The latter type of approach was recently labeled by Horváth (2010) as the *Strong Modularity Hypothesis*, which we characterize here as follows:

Strong Modularity Hypothesis (SMH): No information structure notions – i.e. purely discourse-related notions such as focus (in the sense of focus alternatives of Rooth 1992), contrast (in the sense of explicit focus alternatives), givenness (in the sense of Schwarzschild 1999), or newness (non-givenness) – can be encoded in the grammar as formal features.

The SMH-driven model we have proposed in Fanselow & Lenertová (2010) will have to be elaborated in a number of important respects in the coming four years. We divide this goal into two major subgoals: conducting case-studies serving to further support the SMH and developing a grammatical architecture that is compatible with the core assumptions of the SMH.

#### 3.4.2.1.1. Broadening the empirical scope of our model.

*3.4.2.1.1.1. Languages with different prosodic systems:* Fanselow & Lenertová (2010) have shown that movement to the left periphery in languages such as German or Czech is constrained by cyclic linearization in the sense of Fox & Pesetsky (2005), with early linearization itself being linked to accentuation. In these languages, accentuation precedes movement, and is subject to destressing of given material. The model predicts that languages without clause-internal deaccentuation (Italian) show tighter movement constraints, while languages with accentuation following movement (Hungarian) should be more flexible. Preliminary evidence cited in Fanselow & Lenertová (2010) suggests these predictions are by and large true, but a more systematic investigation will have to be undertaken. The pertinent investigation should also cover languages such as Somali with a putative morphological identification of focus, and languages with different prosodic systems such as tone languages (Chinese, Hausa, Ngamo), pitch accent (Basque) and phrase languages (Hindi). There will be a close interaction with **A5** for the analysis of tone languages.

*3.4.2.1.1.2. Additional types of IS-driven movement:* the SMH has so far been explored mainly with respect to A-bar dependencies (cf. the analysis of “focus-movement” in Fanselow & Lenertová 2010 and contrast in Horváth 2010). We plan to broaden the coverage by including a study about a type of A-movement in Czech, complementing our work on German A-movement (Fanselow, accepted). Kučerová (2007) argues that Czech employs a kind of A-movement which is indirectly induced by a semantic givenness operator (called “G-movement”). Her analysis is incompatible with the SMH since it makes givenness responsible for obligatory syntactic movement. We think that her model is inadequate and that a reanalysis in terms of a prosodically induced movement (Zubizarreta 1998, Arregi 2001, *inter alia*) is called for. There are reasons to believe that Kučerová’s system undergenerates in a number of respects, including conjunctions involving given as well as non-given conjuncts (cf. also Wagner 2006), the position of the main verb, the constituent ordering in the “given area” of the clause, the possibility of non-contrastive new material to c-command given material, and the putative difference between non-specific and specific DPs. Another important point of investigation will be the relation between G-movement and deaccentuation, the latter of which we believe to be available

more generally than predicted by Kučerová. The envisaged outcome of the investigation should be an empirical proof that G-movement is (i) less syntactically restricted and (ii) optional. The theoretical consequence is that the postulation of the G-operator is not justified.

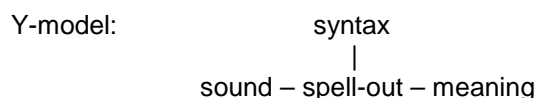
As part of our empirical argument, we intend to conduct at least two acceptability judgement experiments. The first experiment CZ1 will investigate the relative acceptability of structures with G-movement as opposed to structures with deaccenting (D5 will provide expert service). The second experiment CZ2 will investigate the acceptability of structures where non-contrastive new material c-commands given material. We also intend to consult our results with **A7**, with whom we would like to collaborate on related A-type movements in Slavic. Similarly, we will also collaborate with the **B8** researchers, who will be investigating movement into the German right periphery.

*3.4.2.1.1.3. Potential counterexamples in morphology:* The SMH predicts that no language encodes the above-mentioned discourse relations morphologically. Putative information-structure marking morphemes thus need to be reanalyzed, along the lines of some previous studies, such as Zimmermann & Hartmann (**B2**: 2007b), who argued that a putative focus-marker *nee/cee* in Hausa is in fact a marker of exhaustivity (and hence not a focus-marker), or Vermeulen (2007), who showed that *wa* is not a topic marker in Japanese. Concerning morphological marking of givenness, it has been suggested that the West Chadic language Ngamo possesses a background/givenness-marker *ye* (Schuh 2005). In collaboration with **A5**, we will design a questionnaire to test the hypothesis that *ye* is associated with the discourse relation givenness. The preliminary analysis in project **A5** (**A5**: Grubic and Zimmermann, t.a.) takes *ye* to be a definite determiner of events, suggesting that *ye* is not a givenness marker but rather a semantic operator. This study is complemented by two or three further case studies of putative information structure markers, e.g. in Malayalam or in Caucasian languages, the latter on the basis of our long-standing collaboration with Konstantin Kazenin, Moscow. We also plan to consider Féry's (submitted b) idea that putative focus markers may in fact serve a prosodic function of delimiting prosodic phrases (for collaboration with C. Féry, see also 3.4.2.1.3).

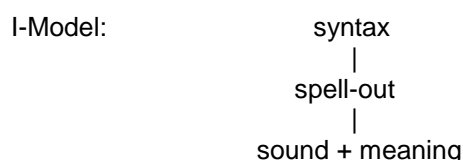
*3.4.2.1.1.4. Potential counterexamples in semantics:* A potential obstacle for the SMH is the so called focus-sensitivity of certain semantic operators, such as *only*, *even*, or *always*. The grammar of these expressions is standardly assumed to make reference to focus, which therefore has to figure in the syntactico-semantic representation, in the form of a feature or at least a pronominal-like index active at LF (Kratzer 1991). In defence of the SMH, we plan to question the feasibility of the assumption that the so-called focus-sensitivity of certain operators is conventionalized (grammaticalized). We plan to follow the recent work of Beaver & Clark (2008), Kadmon & Sevi (2010), Roberts (2010), and A5 (**A5**: Grubic and Zimmermann, t.a.), who argue that the majority (and ideally all) of focus-sensitive expressions associate non-conventionally and hence non-grammatically. Questions of focus association can be addressed with respect to languages that are easily accessible to the members of the project team (German, Spanish, Czech), as well as with respect to some non-Indoeuropean languages, in collaboration with project A5.

#### 3.4.2.1.2. Considerations of grammatical architecture

Syntax has traditionally been thought of as the main module that relates sound with meaning. In the Y-model of grammar (Chomsky 1995; see below), narrow syntax forms the core of sound-meaning mapping, but there are peripheral syntax-like systems (prosody and logical form) which make further adjustments to syntactic representations independently of one another:



That sound and meaning cannot communicate directly is one of the principal reasons for having information structure features in the syntax (see Horvath 2007 for an elaboration). Discourse relations, grounded in the meaning component, clearly exhibit systematic correspondences with sound patterns. The SMH thus requires a shift in the grammatical architecture that allows information to flow between sound and meaning (cf. e.g. Jackendoff 1997, Reinhart 2006). A specific proposal that we would like to work out is “single output syntax”, cf. Richards (1997), here labeled the “I-model”, which is compatible with the SMH and at the same time capable of capturing the sound-meaning correspondence in the domain of information structuring.



In this model, sound and meaning systems access syntactic objects in parallel at spell-out. A full and detailed exposition of this model of grammar and its consequences for the relation between syntax and infor-

mation structure presupposes having answers to a number of fairly general questions. What is the nature of syntactic cycles? What determines the timing of spell-out, i.e. accessing the interfaces? How are sound-meaning mismatches (e.g. inverse scope) to be handled in a model where sound-meaning matching is the default situation? Are they a result of syntactic processes (agreement, formal-feature movement), exceptional lexicalization of lower copies, or simply cases of syntactic underspecification? What is the nature of the sound-meaning interface? Is sound simply a “reflection” of meaning or is there a system of rules by which the two map to one another? Are sound-meaning correspondences part of the generator or do they belong to the OT evaluation model? More generally, how are the rule-based aspects of the envisioned grammatical model (the generator) to be made compatible with the OT evaluation system developed within Goal 2 of this project? Once these questions are answered and the shape of the model is specified, we would like to evaluate its predictions with respect to other, minimalist as well as non-minimalist models.

#### 3.4.2.1.3. Collaboration with visiting professors.

3.4.2.1.3.1. Project A1 has invited Caroline Féry, who has contributed to previous stages of the project, for the spring/summer of 2013. Féry (submitted b) argues that alignment with a prosodic phrase is the cornerstone of a system identifying universal prosodic correlates of focus. The alignment model will be further developed, also on the basis of the new data and insights acquired in the project. The principal goal of the cooperation will be to develop a conception of the syntax-prosody interaction in which the notion of alignment is explored as a more adequate replacement for accent/prominence. Once worked out, this will imply a conceptual broadening of our model of linearization based on accent distribution (Fanselow & Lenertová 2010).

3.4.2.1.4. GLOW 2012: The University of Potsdam will host the GLOW colloquium in 2012, one of the three or four major global conferences in generative syntax. We will have information structure and its consequences for the architecture of grammar as a main theme. This will allow us and the other projects concerned with architectural issues to interact with a top international audience in the context of the questions described in Goal 1. We apply for the funding of the travel costs for the speakers for ten slots and two keynote speakers reserved for the general topic, and two for the related workshops.

### **3.4.2.2 Goal 2: the modelling of gradience and optionality**

Considering the interaction of information structure and syntax, one finds that the data are not categorical (in there being a strict 1-to-1 correspondence between Information Structure factors and their syntactic expression), but rather optional and gradient. This part of the project complements Goal 1 by defining a model of grammar able to account for gradience and optionality, especially regarding the data outputted by Goal 3 (see below); specifically, we will make use of the tools offered by recent developments in Optimality Theory.

#### 3.4.2.2.1. General considerations

As various works have noted, neither standard Minimalism nor standard Optimality theory are adequate to model gradience and optionality (see Scholz and Pullum 2003 and references). Due to this limitation, and following a growing trend (Keller 2000, 2006; Smolensky & Legendre 2006; Jäger & Rosenbach 2006; Hayes & Wilson 2008; Pater 2009; and references), we assume a modification of standard Optimality Theory where each constraint is associated with a *weight* (i.e., a function from constraint violations to positive numbers, with a higher numerical output indicating a more severe violation); the degree of (un)acceptability of an expression can thus be calculated as the sum of the weights of the constraints it violates. A system along these lines enables theorists to quantify the acceptability differences between expressions in a precise way (see especially Keller 2000 for extensive discussion). Instances of such systems, which we collectively refer to as *weighted models*, are *Harmonic Grammar* (HG, Legendre et al 1990, Smolensky and Legendre 2006), *Linear Optimality Theory* (LOT, Keller 2000, 2006), *Stochastic Optimality Theory* (StOT, Boersma 1998, Boersma & Hayes 2001), and *Maximum Entropy* (MaxEnt, Goldwater & Johnson 2003, Jäger 2006). Within this general framework, we will explore the issues detailed in 3.4.2.2.2 through 3.4.2.2.4 below.

#### 3.4.2.2.2. The division of labor between generation and evaluation

A weighted model must be paired with a generative system that provides the expressions to be evaluated. While there is no *a priori* restriction on the kind of system that underlies the generation of candidates, we will adopt the modified minimalist system defined in Goal 1 above, so that both goals can eventually be combined into a coherent architecture. However, combining these two systems is not a trivial matter. To begin with, it is not clear what the position of the evaluation component is within the architecture proposed in Goal 1. Is it placed after the [sound + meaning] stage of the I-model, as usually assumed in work on OT syntax, in which case it can only evaluate completed {structure, sound, meaning} triplets? Does it start at the Spell Out point, in which case it could also affect Spell Out operations like linearization (e.g., the kind of effects discussed in Fanselow and Lenertová 2010)? Or does it run in parallel to the whole derivation (Broekhuis 2008, Heck and Müller 2007), in which case it would also affect narrow syntactic operations? Each alternative partitions the tasks of the generative and the evaluation components differently, with non-trivial consequences.

Similarly, the generative component in Goal 1, in its purest form, outputs just one structure per derivation (as is generally the case with minimalist models), yet the evaluation component is usually assumed to require a multiplicity of candidates as input. One way to resolve this discrepancy is by dropping the latter assumption, so that only one candidate is fed to the evaluator; or, alternatively, we can modify the generator so that it creates multiple candidates. As before, each option attributes different responsibilities to each component, and the differences that arise are not trivial. The answer to these questions will be based on a close inspection of the data produced by Goal 3 (and, to a lesser extent, Goal 1 and other SFB 632 projects concerned with the syntax-information structure interface, viz, **A5**, **A6**, **A7** and **B8**). More specifically, we will map out the predictions made by each different conception of the generation/evaluation division of labor and compare them against the data produced by the empirical studies.

#### 3.4.2.2.3. The relation between acceptability ratings and corpus frequencies

Different weighted models make use of different sorts of data in order to determine the weight of each constraint. LOT requires acceptability judgements gathered through standard elicitation techniques, whereas StOT and MaxEnt require frequency data gathered from corpus analysis. Each method has its advantages and disadvantages. StOT and MaxEnt are preferable on the grounds that they model human language acquisition more closely – i.e., human infants do not have access to the acceptability judgements of adults (the data required by LOT), but they are sensitive to the frequency of expressions in adult language (Yang 2002 and references). At the same time, StOT and MaxEnt also rely on the assumption that there is a well-defined positive correlation between corpus frequency and relative acceptability (cf. Jäger and Rosenbach 2006, Hayes and Wilson 2008). This assumption, however, has been challenged (Culy 1998, Keller and Asudeh 2002), and its current status is unclear (but see Häussler & Bader 2010 for some advances).

This is an important issue to resolve, as the general shape of the evaluation model depends on the kind of data that are required to define constraint weights (i.e., if it turns out that there is no generalized function from corpus frequencies to grammaticality judgements, then the standard StOT/MaxEnt weighting methods will have to be discarded). In order to arrive at an answer, we plan a collaboration with project A6 to compare corpus frequencies and grammaticality judgements in a specific domain – namely, the effect of information structure in the linearization of double object constructions in Germanic and Romance languages. This domain has been chosen for various reasons, viz., the fact that there already exists some work on this area (Bader 2010, Häussler and Bader 2010) that we can use as a basis, the fact that the relevant constructions are relatively frequent and easy to identify in corpora, and the fact that it involves a small number of variables (i.e., the relative position of two DPs and the placement of accent), which makes the comparison quite manageable. In later phases, the utterance corpus described in 3.4.2.3.3 will also be analysed.

#### 3.4.2.2.4. Development of a constraint-weighting method

A weighted model must be combined with an algorithm to calculate constraint weights; ideally, this algorithm should reflect the way in which (human) learners come to know the correct weights. However, existing algorithms, even though they are useful for the theorist, fail to satisfy this desideratum. Some of them (Keller 2000, Hayes & Wilson 2008) work offline – i.e., they require a relatively large body of data to be gathered before the weighting process can begin, and no additional data can be added during the process (in fact, the addition of even a single datum requires the whole process to restart from scratch). This is unrealistic from a human learner's perspective, where data gathering and constraint weighting proceed in unison. On the other hand, algorithms that work online require the learner to have an unrealistic degree of advance knowledge (e.g., Jäger 2006 requires the learner to know the actual target weights to be attained). In this subpart of the project, we will aim to circumvent these problems by developing a constraint-weighting method that operates both (i) online, and (ii) on the basis of naturally available evidence only. Since there are already algorithms satisfying either (i) or (ii), but not both, our first goal will be to determine whether any of these can be extended to cover both desiderata. Should this strategy not yield any results, we will attempt to define a new algorithm that performs well on both counts.

### **3.4.2.3 Goal 3: The establishment of the empirical basis for modelling.**

We plan a series of experiments studying the influence of various pragmatic factors on word order. In order to keep the amount of experiments manageable, our program must be confined to focus placement. The experiments serve two functions. First, they constitute the database for the modelling subproject. Second, they will shed light on the relative importance of information structure in the narrow sense and other aspects of context and the communicative situation with respect to the choice of particular word orders.

#### 3.4.2.3.1. Factors to be Studied

Basic concepts of information structure such as focus rarely force the movement of an XP to the left periphery. Rather, it seems that they have an effect on word order only in combination with further criteria such as contrast (Rizzi 1997) or exhaustivity (É. Kiss 1998) also belonging to the information structure of a clause. However, the preparation of a topic shift in the next clause (**C1**: Hörnig et al, Ms.) and the mere presence of



a POSET-relation (**C1**: Weskott et al., accepted) motivate the leftward movement of XP, too, although they do not characterize the information structure status of XP within its clause. In other words there are context effects beyond information structure, and we want to assess the role of these two types of context properties

The first group of factors to be studied consists of contrast/correction, exhaustivity and predictability that refer to properties a phrase has relative to the content and information structure of the clause it occurs in. EXP1 will assess the effect of contrast as a function of the explicitness of the alternatives (cf. Repp 2010a), including explicit correction and mere POSET relations (cf. Prince 1998), while the effects of predicability and exhaustivity will be measured in EXP2. EXP3 and EXP4 study the interaction between these factors, which can sometimes go in unexpected directions. Thus, Skopeteas & Fanselow (**D2**: submitted) found that low predictability reduces the likelihood of an exhaustive interpretation of a marked word order. A context to be judged for acceptability could look like the following with a manipulation of predictability.

- A(i): Heute morgen war Fritz am Bodensee beim Angeln  
 B: Und was hat er gefangen?  
 A(ii): Einen Fisch/Einen Karpfen/einen Reisewecker hat er gefangen

The second class of factors involves text organization. A first subclass relates to text coherence, including the preparation of a topic shift (cf. **C1**: Hörning et al, Ms.) and the signalling of intertextual relations. E.g., we found that a sentence with marked word order may be more acceptable than its SVO counterpart when the two sentences stand in an adversative relation, but that effect is minimized or disappears when adversativity is explicitly signalled by *aber* 'but'. Ideally, two such relations and topic shift can be studied within EXP 5, while EXP6 is concerned with the interaction of topic shift with contrast and exhaustivity.

The choice of a marked order may serve to add positive or negative emotional connotation to an utterance (the marked sentence *Windeln hab ich gewechselt* as a response to "what did you do yesterday" comes with a flavor of joy or despair). A marked order may also be chosen in order to guarantee a high degree of hearer attention. We plan to study expressive constellations in combination with the exhaustivity-contrast constellation and a predicability manipulation (EXP7). Attention control should be combined with a predictability manipulation (EXP8).

These 8 experiments involve a mean of 5 pragmatic distinctions. These combine with the difference between marked (object initial) and unmarked word order, and with the gross pragmatic difference between subject, object and wide focus. Each experiment will thus involve  $5 \times 6 = 30$  factors. If we use 4 items per condition, we arrive at 120 experimental items, with the material being distributed on 2 subexperiments each.

#### 3.4.2.3.2. Languages to be Studied

We want to build our modelling on a pair of relatively different (German, Spanish), and relatively similar (European and American Spanish) languages; the choice of a Romance language is motivated by the stronger impact of contrast and correction on focus movement. The European Spanish studies will be carried out in Barcelona, for American Spanish we collaborate with Diego Quesada, San José. We also would like to run the contrast study for Italian and Finnish, and the exhaustivity experiment for Hungarian, because of the import ascribed to these factors for the respective languages in the literature. In collaboration with project A5, we plan to run part of the experiments for Hausa (suspected subject/object asymmetry in focus marking) and Ngamo (use of morphological background markers in the expression of focus-background partition).

#### 3.4.2.3.3. Method

Written questionnaire studies in which participants rate the acceptability of a sentence in context will constitute the bulk of our experiments (EXP1-6). While the Ngamo experiments will be carried out with the help of project A5 during a field visit, the Hausa experiments will have to be internet-based, also in close collaboration with project A5. In EXP 7, participants will rate the emotional quality of texts that differ in terms of a word order manipulation in one sentence, in addition to their acceptability. A good design for control of attention experiments (EXP 8) has participants carry out a task linked to the comprehension of texts that involve the word order manipulation in question.

For German, EXP1-6 will be carried out as production experiments, too. This allows us to build up a controlled utterance corpus for word order choice by which we can systematically relate acceptability to corpus frequency. To the best of our knowledge, such a data set does not exist yet. A context will be presented to the participants, who will be asked to produce a marked/unmarked structure in this context. The method to be employed could be a forced choice experiment, a sentence completion task (e.g., Fanselow & Weskott 2010) or guided spontaneous production. A decision will be taken after a pilot study.

#### 3.4.2.3.4. Demands

Close to identical material will have to be constructed for EXP1- EXP8 in German and Spanish, and about eight experiments for the other languages. There are 16 (acceptability) plus 12 (production) = 28 subexperi-

ments in German, 16 in Spanish (run twice in Europe and America) plus at most 16 subexperiments (2x8) for the other languages. We thus apply for the funding of 76 subexperiments with 24 subjects each.

### 3.4.3 Timetable

We divide the four years of this research phase into eight semesters:

Semester 1 07/2011 – 12/2011	- Investigating the relation between givenness and syntax (in Czech) - EXP CZ1 - EXP 1 & 2 (German: acceptability & production; Spanish: European & American)
Semester 2 01/2012 – 06/2012	- Developing a prosodically-based analysis of “givenness” movement - EXP CZ2 - Collaboration with <b>A6</b> (comparison acceptability ratings vs corpus frequencies). - EXP 3 & 4 (Ger: acceptability & production, Spa)
Semester 3 07/2012 – 12/2012	- Reserach on putative morphological markers of information structure - EXP Hausa/Ngamo (collaboration with <b>A5</b> )
Semester 4 01/2013 – 06/2013	- Investigation and reanalysis of conventional (grammatical) association with focus - EXP 5 & 6 (Ger: acceptability & production, Spa)
Semester 5 07/2013 – 12/2013	- Exploring the predictions of Fanselow & Lenertová (2010) for different prosodic systems, developing an amended version of the theory of cyclic linearization. - EXP Italian, Finnish, Hungarian
Semester 6 01/2014 – 06/2014	- Development of a grammatical architecture compatible with discourse free syntax. - Research on the division of labor between the generator and the evaluation module. - EXP 7 (Ger, Spa)
Semester 7 07/2014 – 12/2014	- Development of a grammatical architecture compatible with discourse free syntax. - Integrating the generator module with the evaluation module. - EXP8 (Ger, Spa)
Semester 8 01/2015 – 06/2015	- Evaluation of predictions with respect to other (non-minimalist) models. - Integration of experimental results

### 3.8 Weitere, (noch) unveröffentlichte Projektarbeiten (s. beiliegende CD)

Féry, C. (submitted a). German Sentence Accents and Embedded Prosodic Phrases. Re-submitted to *Lingua*.

Féry, C. (submitted b). Focus realization as Prosodic Alignment. Ms. Universität Frankfurt/M.

Ishihara, S. (submitted a). Japanese focus prosody revisited: Freeing focus from prosodic phrasing. Re-submitted to *Lingua*.

Ishihara, S. (submitted b). Japanese Downstep Revisited. Ms. Universität Frankfurt/M.

Ishihara, S., B. Surányi & F. Schubö (submitted). Syntax-Prosody Mapping, Topic-Comment Structure, and Stress-Focus Correspondence in Hungarian.